Task 3. Visualize a data story (35%)

Renewable Energy

studentNumber

Data:

For this exercise, use the `renewableEnergyProduction.csv` file. This dataset contains yearly information on renewable energy production of many countries in the world and contains the following variables:

- `Entity`: Country or region name
- `Code`: Country or region code. Country codes are the official codes, and are also used in other datasets, e.g. it is similar to the variable `adm0_a3` of the data in the `rnaturalearth` package.
- `Year`: Year
- `Wind`: Electricity from wind, in TWh
- `Hydropower` = Electricity from hydropower, in TWh
- `Solar`: Electricity from solar, inTWh
- `Other`: Electricity from other renewables including bioenergy, in TWh
- `Total`: Electricity from the renewables combined (i.e. total of `Wind`, `Hydropower`, `Solar`, and `Other`), in TWh
- `Share_Percentage`: The share of electricity production that comes from renewable technologies, percentage

Assignment:

Assume the role of an analyst at an institute that promotes the use of renewable energy. Your task is to analyze global renewable energy production trends. Specifically, to identify and report on disparities in renewable energy production around the world, to guide the institute's focus on specific regions and energy resources.

Objectives:

- Make a data story for the above scenario.

- Produce 2-3 visualizations that show:
  - The global distribution of renewable energy production.
  - The trends in different types of renewable energy production over time.

Note: dataset has been pre-cleaned, but additional data wrangling might be needed, depending on the information you want to show.

Report and wishlist:

- **Report:** A brief explanation of your choices, e.g. why you chose the particular type of graph and its design (colors, etc), and why you added any elements (if applicable, e.g. labels and other elements included to add extra information).
- **Wishlist:** Provide a list of the (max) 5 most important things that you would have liked to change/add to your plot if you would have had more time, and briefly explain why. Be specific in your description of changes, e.g. do not write "I would change the color.", but describe in detail what type of color scale and which colors you would have chosen.

Load data:

```
# Load packages
library(tidyverse)
library(foreign)

# Import data:
dataRenEnergy_Prod <- read_csv("renewableEnergyProduction.csv")
```

Your data story - plots

```
# your code here
summary(dataRenEnergy_Prod)
```

```
##     Entity              Code                Year           Wind
##  Length:4756        Length:4756        Min.   :1985   Min.   :  0.00
##  Class :character   Class :character   1st Qu.:2001   1st Qu.:  0.00
##  Mode  :character   Mode  :character   Median :2008   Median :  0.00
##                                        Mean   :2007   Mean   :  3.17
##                                        3rd Qu.:2015   3rd Qu.:  0.10
##                                        Max.   :2022   Max.   :762.70
##
##    Hydropower          Solar              Other             Total
##  Min.   :   0.00   Min.   :  0.000   Min.   :  0.000   Min.   :   0.00
##  1st Qu.:   0.17   1st Qu.:  0.000   1st Qu.:  0.000   1st Qu.:   0.43
##  Median :   2.79   Median :  0.000   Median :  0.021   Median :   3.77
##  Mean   :  22.65   Mean   :  1.276   Mean   :  2.501   Mean   :  29.73
##  3rd Qu.:  12.22   3rd Qu.:  0.020   3rd Qu.:  0.820   3rd Qu.:  16.17
##  Max.   :1321.71   Max.   :427.720   Max.   :176.629   Max.   :2670.18
##  NA's   :45                          NA's   :15        NA's   :60
##  Share_Percentage
##  Min.   :  0.000
##  1st Qu.:  5.354
##  Median : 22.096
##  Mean   : 34.237
##  3rd Qu.: 60.191
##  Max.   :100.000
##
```

```
cleaned <- na.omit(dataRenEnergy_Prod)
```

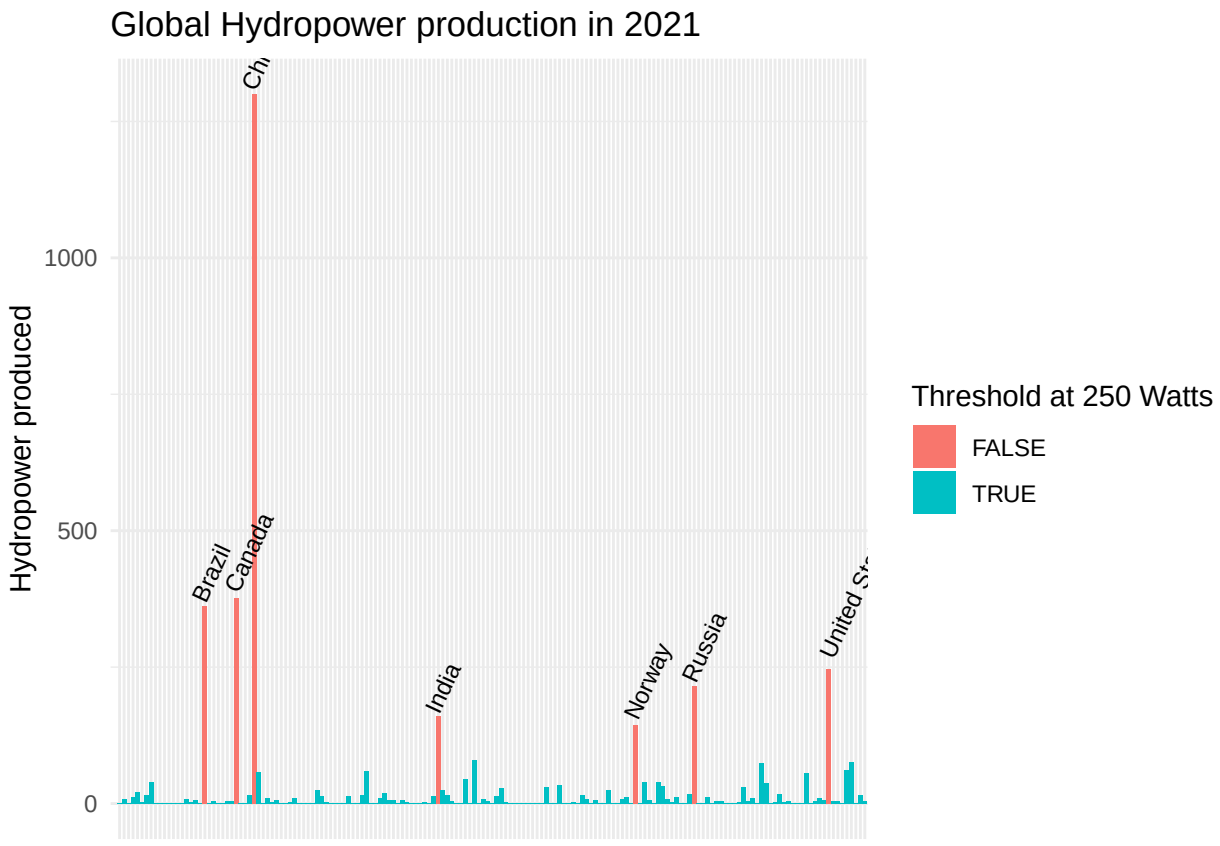
```
unique(cleaned$Year)
```

```
##  [1] 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014
## [16] 2015 2016 2017 2018 2019 2020 2021 1990 1991 1992 1993 1994 1995 1996 1997
## [31] 1998 1999 2022 1985 1986 1987 1988 1989
```

```
#We then pick the latest year 2021


data_2021 <- cleaned[which(cleaned$Year == "2021"),]

data_2021_more_than <- cleaned[which(cleaned$Hydropower > 10),]

threshold <- 100
```

```r
# Create a new column indicating whether the estimate is below the threshold
data_2021_final <- data_2021 %>%
  mutate(below_h_threshold = Hydropower < threshold)


ggplot(data = data_2021_final,
       mapping = aes(
         x = Entity,
         y = Hydropower,
         fill = below_h_threshold

       )) + geom_bar(stat="summary", fun="mean", nbins=5) + theme_minimal() +
theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) + labs(x = "Countries",
                                             y ="Hydropower produced",
                                             fill = "Threshold at 250 Watts") +
  geom_text(aes(label = ifelse(!below_h_threshold, Entity, "")),
            hjust = -0.1, color = "black", size = 3, angle = 65, nudge_x = -2) +
  ggtitle("Global Hydropower production in 2021")
```

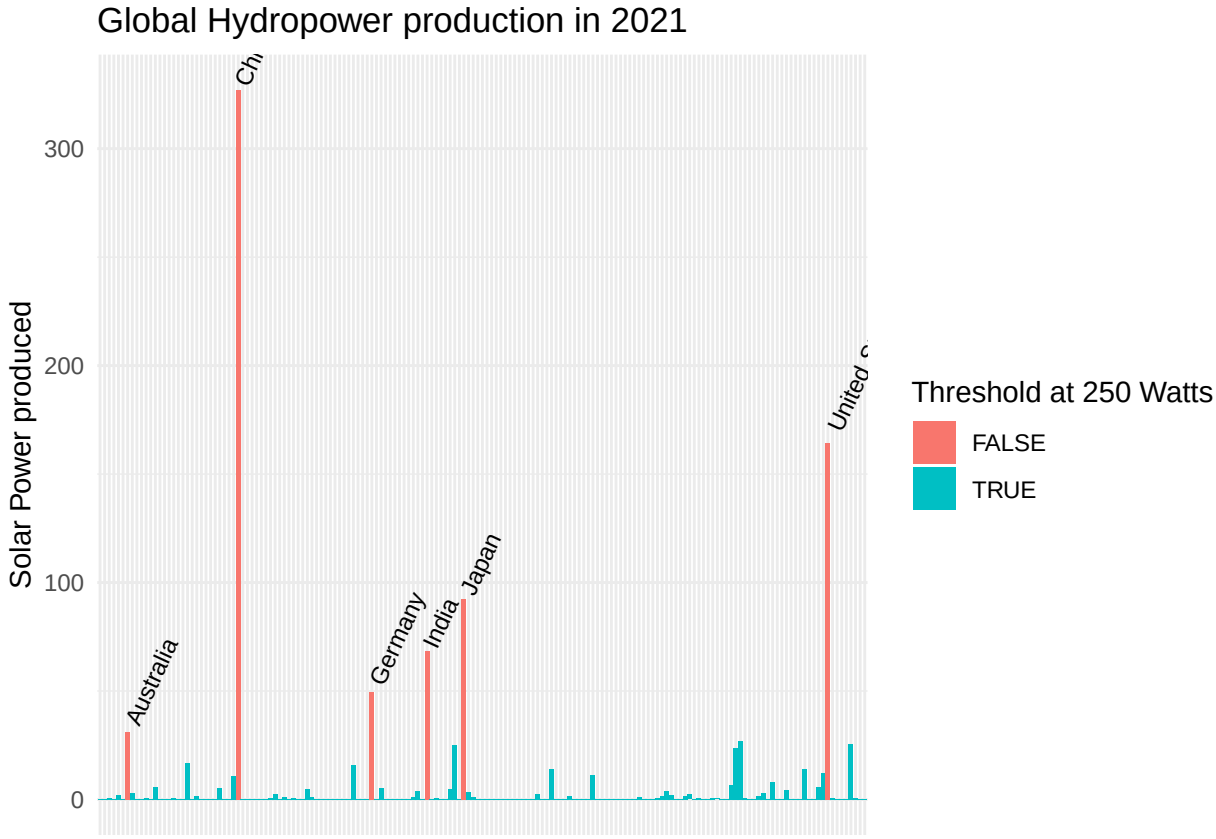
```
## Warning in geom_bar(stat = "summary", fun = "mean", nbins = 5): Ignoring
## unknown parameters: `nbins`
```




```r
threshold = 30

data_2021_final <- data_2021 %>%
```

```
  mutate(below_s_threshold = Solar < threshold)

ggplot(data = data_2021_final,
       mapping = aes(
         x = Entity,
         y = Solar,
         fill = below_s_threshold

       )) + geom_bar(stat="summary", fun="mean", nbins=5) + theme_minimal() +
theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) + labs(x = "Countries",
                                             y ="Solar Power produced",
                                             fill = "Threshold at 250 Watts") +
  geom_text(aes(label = ifelse(!below_s_threshold, Entity, "")),
            hjust = -0.1, color = "black", size = 3, angle = 65) +
  ggtitle("Global Hydropower production in 2021")
```

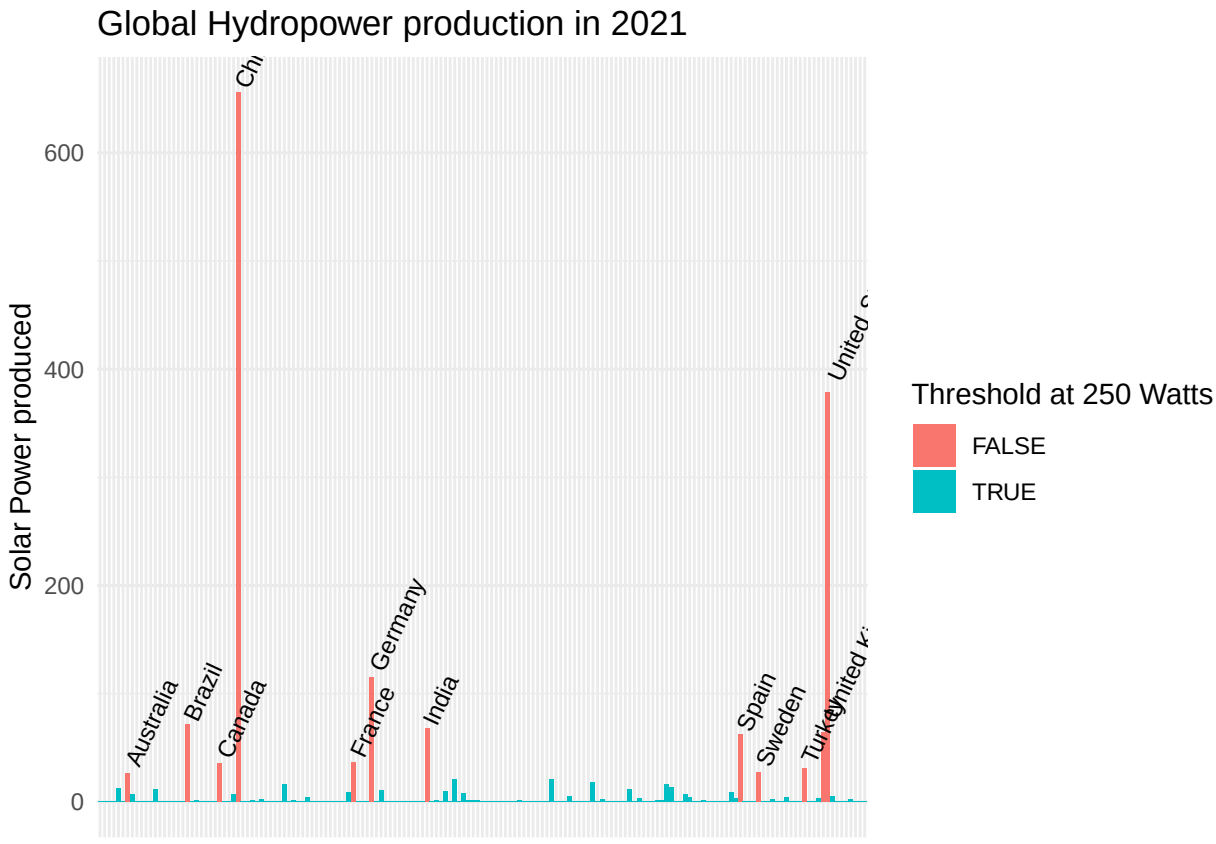
```
## Warning in geom_bar(stat = "summary", fun = "mean", nbins = 5): Ignoring
## unknown parameters: `nbins`
```




```
threshold = 25

data_2021_final <- data_2021 %>%
  mutate(below_w_threshold = Wind < threshold)

ggplot(data = data_2021_final,
```

```
       mapping = aes(
         x = Entity,
         y = Wind,
         fill = below_w_threshold

       )) + geom_bar(stat="summary", fun="mean", nbins=5) + theme_minimal() +
theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) + labs(x = "Countries",
                                             y ="Solar Power produced",
                                             fill = "Threshold at 250 Watts") +
  geom_text(aes(label = ifelse(!below_w_threshold, Entity, "")),
            hjust = -0.1, color = "black", size = 3, angle = 65) +
  ggtitle("Global Hydropower production in 2021")
```

```
## Warning in geom_bar(stat = "summary", fun = "mean", nbins = 5): Ignoring
## unknown parameters: `nbins`
```

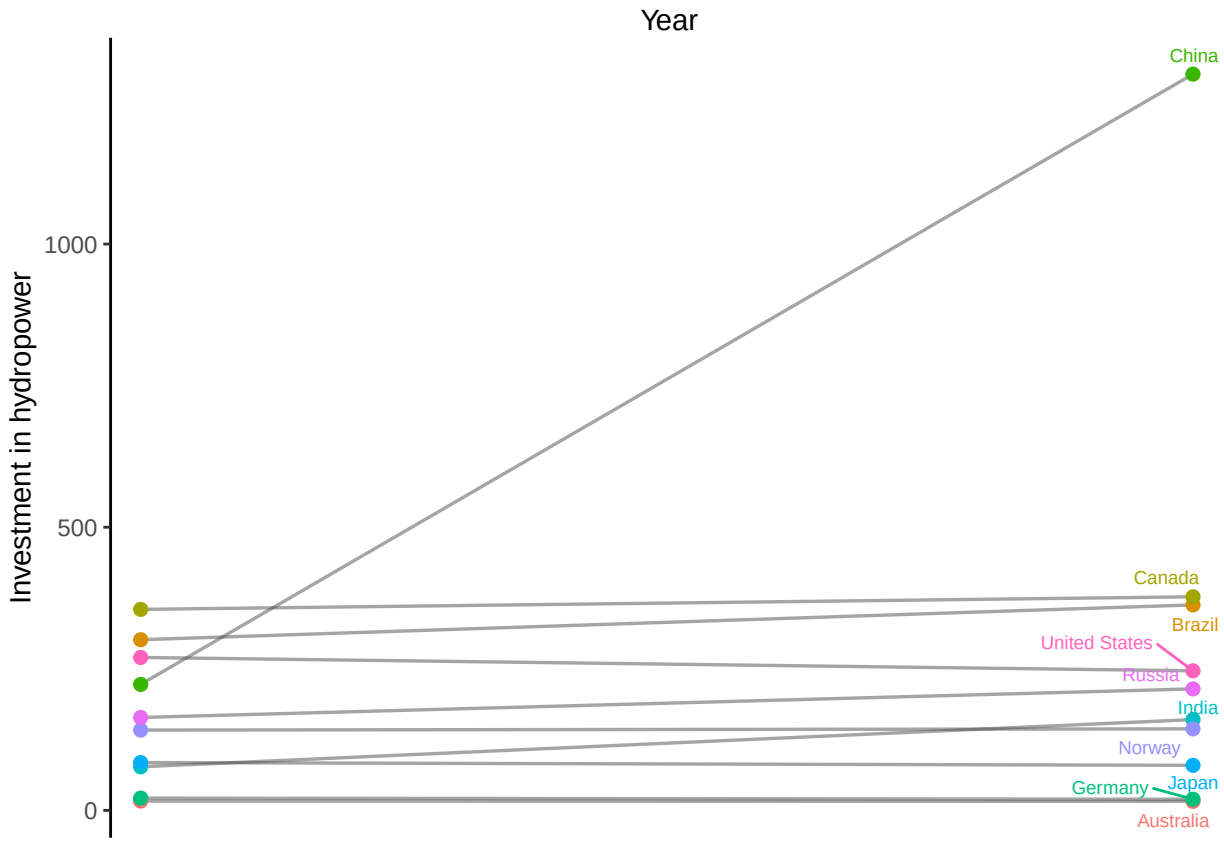



```
library(ggrepel)



data_2000_2021 <- cleaned[which(cleaned$Year %in% c("2021","2000")),]

#data_2021_more_than <- cleaned[which(cleaned$Hydropower > 10),]

threshold <- 250
```

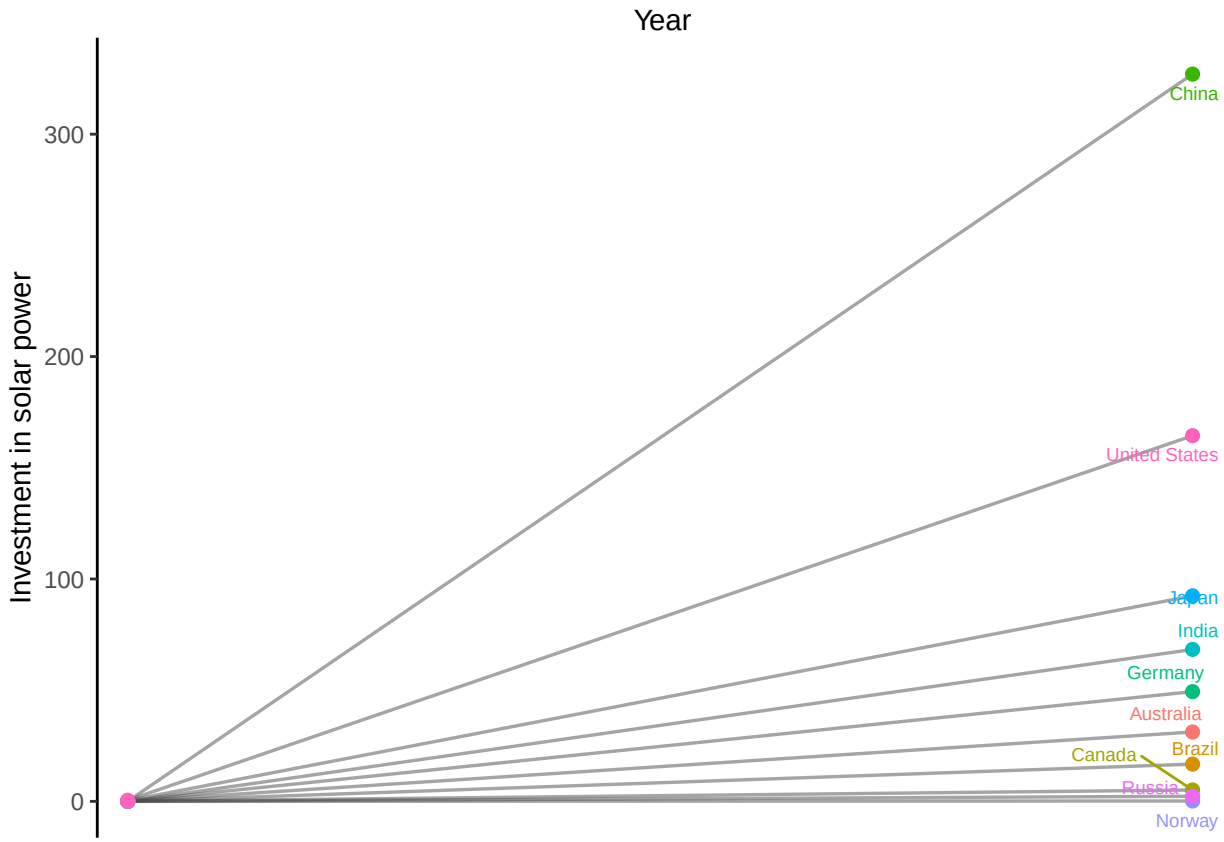
```r
selected_countries <- data_2000_2021[which(data_2000_2021$Entity %in% c("Australia",
                                                          "Germany",
                                                         "India",
                                                         "Japan",
                                                         "China",
                                                         "Norway",
                                                         "Russia",
                                                         "Brazil",
                                                         "Canada",
                                                         "United States")),]
splot3 <- ggplot(data = selected_countries, aes(
      x = Year,
      y = Hydropower,
      group = Entity,
      color = Entity)) +
      geom_line(size = 0.6, alpha = 0.5, colour = "grey30") +
      geom_point( size = 2) +

  geom_text_repel(data = subset(selected_countries, Year == "2021"),
  aes(label = Entity),
  size = 2.5,
    nudge_x = 0.1) +

labs(y = "Investment in hydropower") +
scale_x_discrete(position = "top", labels = c(
      "2000" = "2000",
      "2021" = "2021"
    )) + theme_classic() + theme(legend.position="None")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

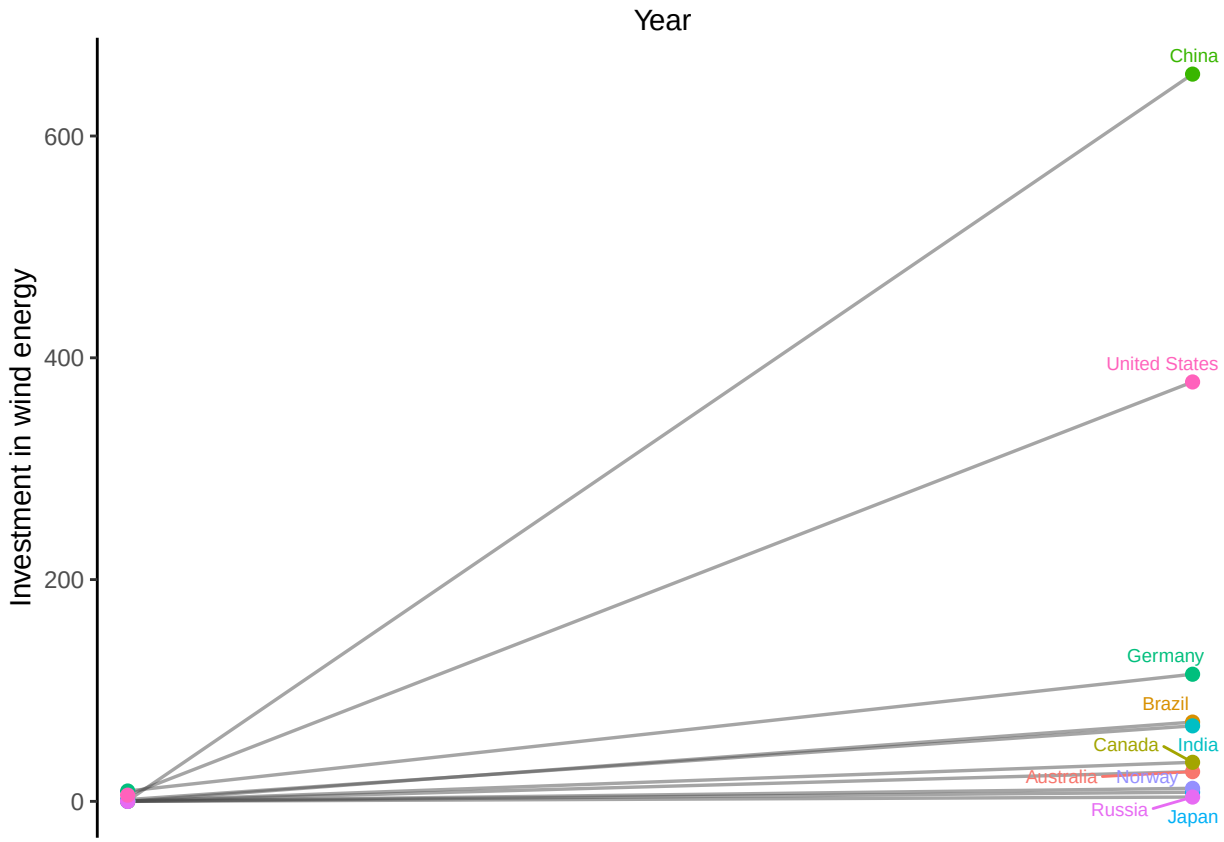
```r
splot3
```




```
splot2 <- ggplot(data = selected_countries, aes(
      x = Year,
      y = Solar,
      group = Entity,
      color = Entity)) +
      geom_line(size = 0.6, alpha = 0.5, colour = "grey30") +
      geom_point( size = 2) +

  geom_text_repel(data = subset(selected_countries, Year == "2021"),
  aes(label = Entity),
  size = 2.5,
    nudge_x = 0.1) +

labs(y = "Investment in solar power") +
scale_x_discrete(position = "top", labels = c(
      "2000" = "2000",
      "2021" = "2021"
    )) + theme_classic() + theme(legend.position="None")

splot2
```

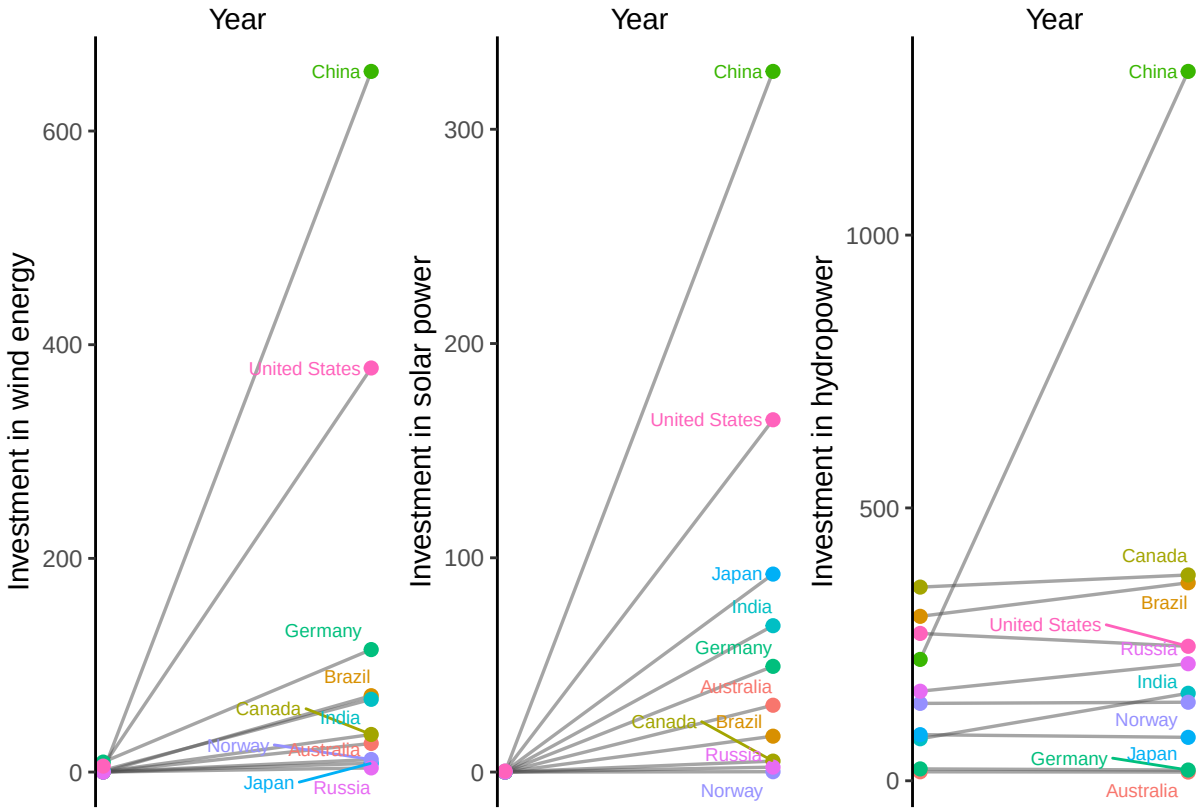



```
splot1 <- ggplot(data = selected_countries, aes(
      x = Year,
      y = Wind,
      group = Entity,
      color = Entity)) +
      geom_line(size = 0.6, alpha = 0.5, colour = "grey30") +
      geom_point( size = 2) +

  geom_text_repel(data = subset(selected_countries, Year == "2021"),
  aes(label = Entity),
  size = 2.5,
    nudge_x = 0.1) +
labs(y = "Investment in wind energy") +
scale_x_discrete(position = "top", labels = c(
      "2000" = "2000",
      "2021" = "2021"
    )) + theme_classic() + theme(legend.position="None")


splot1
```



Choice 1

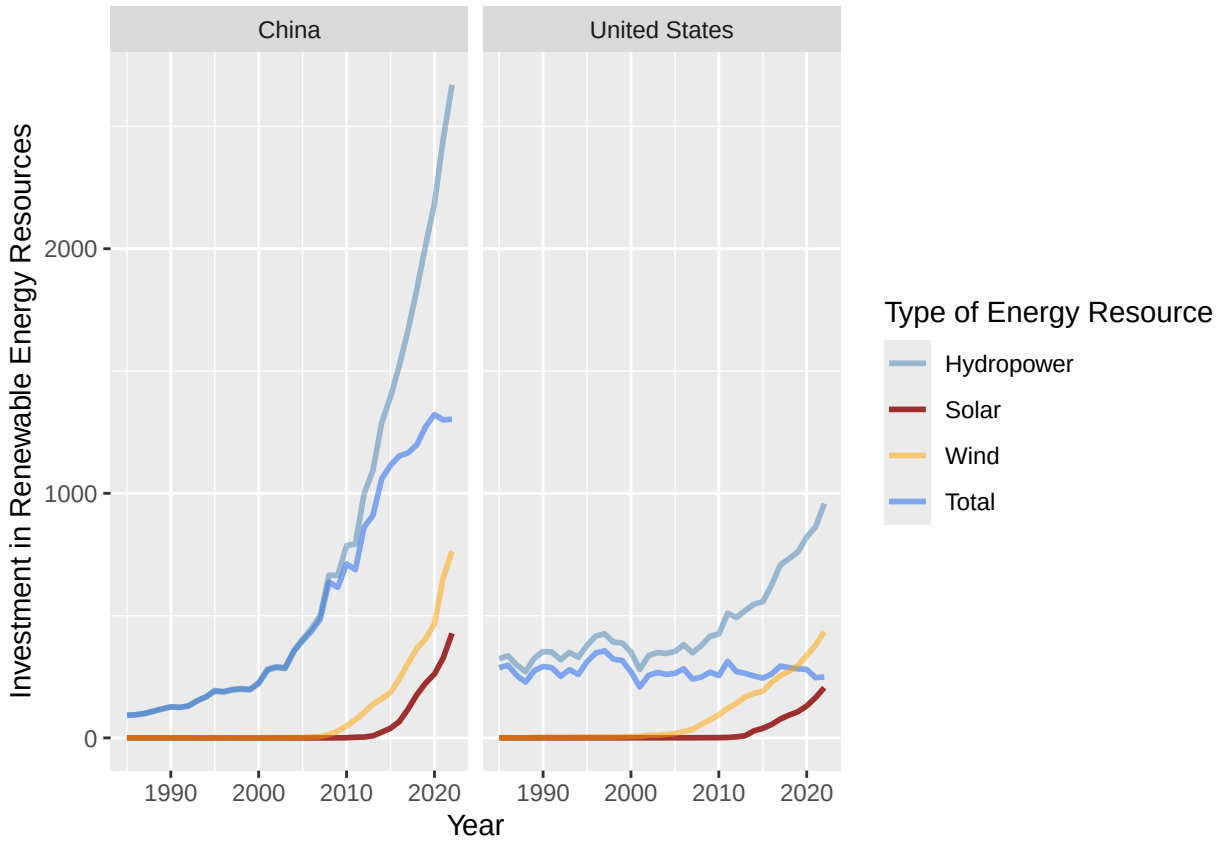
```
library(ggpubr)

#final_plot <-splot1 + splot2 + splot3
#ggarrange(splot1,splot2, splot3 + rremove("x.text"),
 #          labels = c("A", "B", "C"),
 #          ncol = 2, nrow = 2)

library(patchwork)
final_plot <-splot1 + splot2 + splot3
final_plot
```




```
selected_countries_2 <- cleaned[which(cleaned$Entity %in% c("Australia",
                                                           "Germany",
                                                           "India",
                                                           "Japan",
                                                           "China",
                                                           "Norway",
                                                           "Russia",
                                                           "Brazil",
                                                           "Canada")),]
```

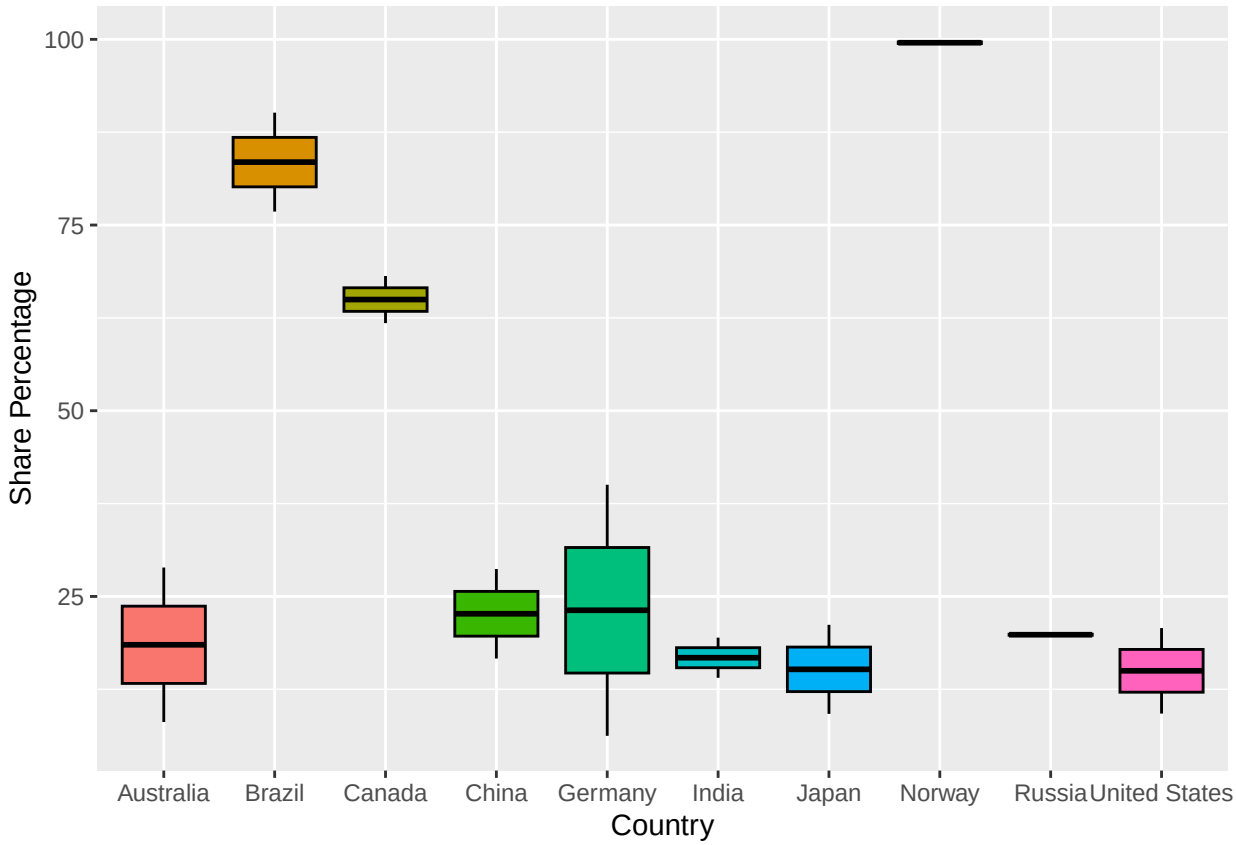
Choice 2

```
cus <- cleaned[which(cleaned$Entity %in% c("China", "United States")),]

ggplot(cus, aes(x=Year)) +
  geom_line(aes(y = Hydropower, color = "steelblue"), alpha=0.8, linetype="solid", linewidth=1) +
  geom_line(aes(y = Solar, color="darkred"), alpha=0.8, linewidth=1 ) +
  geom_line(aes(y = Wind,  color="orange"), alpha=0.5, linewidth=1) +
  geom_line(aes(y = Total,  color="cornflowerblue"), alpha=0.5, linewidth=1) +
  labs(color="Type of Energy Resource") +
  facet_grid(cols=vars(Entity)) +
  scale_color_manual(labels = c("Hydropower", "Solar","Wind", "Total"),
                     values = c("steelblue", "darkred", "orange", "cornflowerblue")) + labs(y="Investmer
```



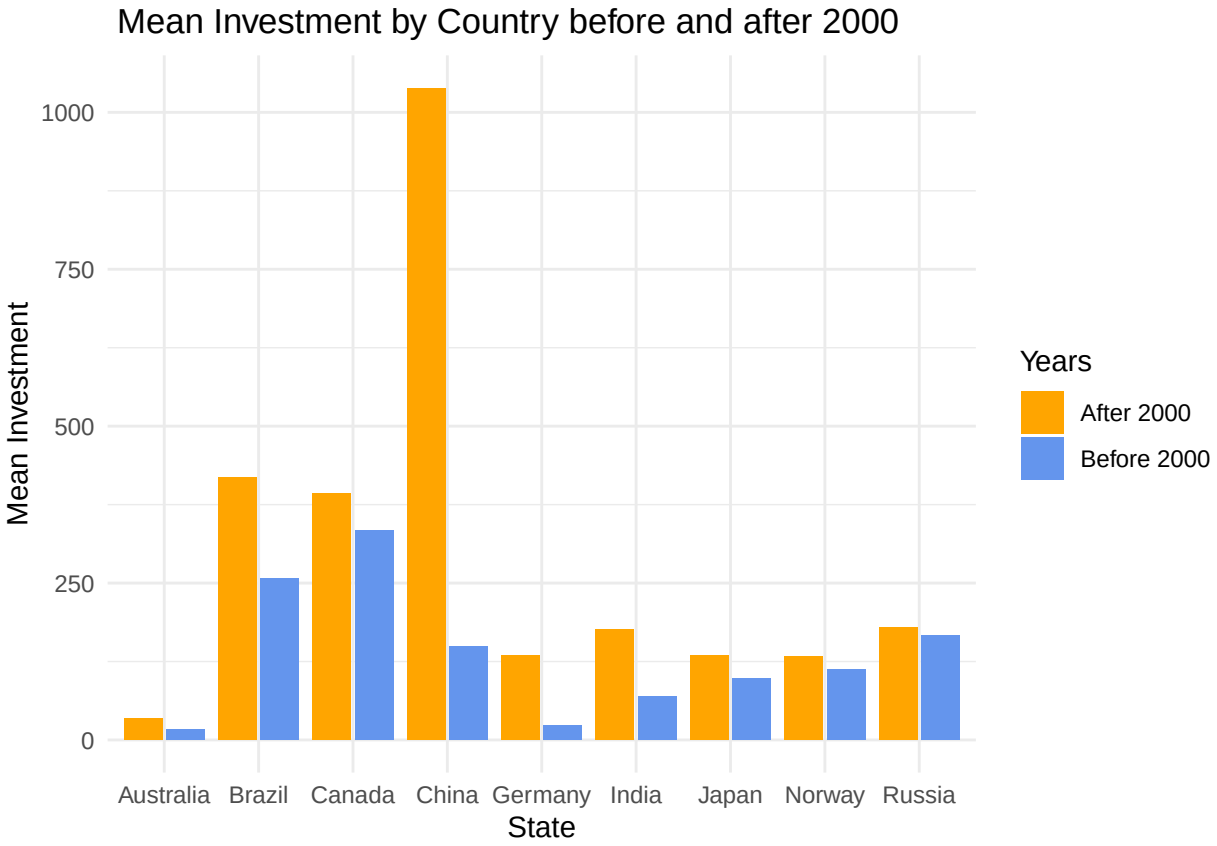
Choice 3

```
ggplot(data = selected_countries,
       mapping = aes(
         x = Entity,
         y = Share_Percentage,
         fill = Entity
       )) + geom_boxplot(color="black") + theme(legend.position="none") + labs(x="Country", y="Share Pe
```



###Choice 4

```
threshold = 2000
year_comp <- cleaned %>%
  mutate(before_2000 = Year %in% c("1999", "1998", "1997", "1996", "1995",
                                   "1994", "1993", "1992", "1991", "1990", "1989", "1987",
                                   "1986", "1985"))

year_comp_se <- year_comp[which(year_comp$Entity %in% c("Australia",
                                                        "Germany",
                                                       "India",
                                                       "Japan",
                                                       "China",
                                                       "Norway",
                                                       "Russia",
                                                       "Brazil",
                                                       "Canada")),]


# Plot the data
ggplot(data = year_comp_se, aes(x = Entity, y = Total, fill = before_2000)) +
  stat_summary(fun = "mean", geom = "bar", position = "dodge2") +
  scale_fill_manual(values = c("FALSE" = "orange", "TRUE" = "cornflowerblue"),
                    labels = c( "FALSE" = "After 2000", "TRUE" = "Before 2000")) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "State", y = "Mean Investment", fill = "Years",
       title = " Mean Investment by Country before and after 2000") + theme_minimal()
```



Report - briefly explain your choices

- choice 1
- choice 2
- choice 3

For choice 1 I chose a combined spaghetti plot to show the differences in investment from 2000 to 2021 of different countries in solar, wind and hydropower. This is because climate change activism has really flourished since the 2000's and it would be exciting to see how far we have come in this century. Sadly the breadth of data made it such that not all countries could be represented properly, thus I used an initial set of graphs and summaries to deduce which countries have more prominent investments in these ventures. For this venture I chose spaghetti plots since they are helpful in representing changes over a passage of time. Additionally, they are handy when dealing with a large number of categorical variables (like Entity or countries in usual cases) as each variable value is represented with a point which does not take much space and leaves little chance of overlap and confusion.

For choice 2 I narrowed my focus on China and the United States in particular. This is because China and the US ha starkly increased their investments in solar, wind, hydropower and renewable energy resources in general since the 2000's. It would be nice to see who got there first. For this venture I chose a facet grid plot with two subplots, one for China and one for the US. This way I managed to save on the extra colors and lines I would have had to add if I put theme in one graph. From our data it is clear that China got there first with booming investments in renewable energy resources post 2010.

For choice 3 I plotted the share percentage for the top 10 countries from the years 2000 to 2021 using a box plot. This is because I believe that it is important to highlight the share of electricity production that comes from renewable resources for each country. I chose a boxplot because a boxplot, seemingly simple can actually plot variations in data very well, especially when there is multiple values to choose from. It does that by having a tail that can be extended to show variance. We can see that Brazil, Canada and Norway's

electricity production is primarily sourced from renewable energy, while countries like Russia, United States and China are seemingly falling behind.

For choice 4, I plotted the average total investment in renewable energy before and after 2000. For this I used a simple bar graph as it can help with portraying differences over time (with time as a categorical variable in this case) while conserving space and color. As seen from the graphs it is highlighted that total investments in renewable energy increased post 2000.

Wishlist - what would you have liked to change/add?

1.I wish i could have fixed some errors in year plotting in the spaghetti plot as I cannot get it to display the years underneath the points. Furthermore, I am not successful in flipping the "before 2000" and "after 2000"
2.
3. 4. 5.